Internet Engineering Task Force
INTERNET DRAFT
draft-floyd-tcp-slowstart-00.txt
Sally Floyd
ICSI
June, 2002

Limited Slow-Start for TCP with Large Congestion Windows

Status of this Memo

Abstract

This note proposes a modification for TCP's slow-start for use with TCP connections with large congestion windows. For TCP connections that are able to use congestion windows of thousands (or tens of thousands) of MSS-sized segments (for MSS the sender's MAXIMUM SEGMENT SIZE), the current slow-start procedure can result in increasing the congestion window by thousands of segments in a single round-trip time. Such an increase can easily result in thousands of packets being dropped in one round-trip time. This is often counter-productive for the TCP flow itself, and is also hard on the rest of the traffic sharing the congested link. This note proposes Limited Slow-Start, limiting the number of segments by which the congestion window is increased for one window of data during slow-start, in order to improve performance for TCP connections with large congestion windows.

1. Introduction

This note proposes a modification for TCP's slow-start for use with TCP connections with large congestion windows. For TCP connections that are able to use congestion windows of thousands (or tens of thousands) of MSS-sized segments (for MSS the sender's MAXIMUM SEGMENT SIZE), the current slow-start procedure can result in increasing the congestion window by thousands of segments in a single round-trip time. Such an increase can easily result in thousands of packets being dropped in one round-trip time. This is often counter-productive for the TCP flow itself, and is also hard on the rest of the traffic sharing the congested link. This note proposes Limited Slow-Start, limiting the number of segments by which the congestion window is increased for one window of data during slow-start, in order to improve performance for TCP connections with large congestion windows.

When slow-start results in a large increase in the congestion window in one round-trip time, the subsequent drop of a large number of packets in the network can result in unnecessary retransmit timeouts for the TCP connection. The TCP connection could end up in congestion avoidance phase with a very small congestion window, and could take a large number of round-trip times to recover its old congestion window. This poor performance is illustrated in [F02].

2. The Proposal for Limited Slow-Start

Limited Slow-Start introduces a parameter, "max_ssthresh", and the slow-start is only modified for values of the congestion window "cwnd" greater than "max_ssthresh". That is, for:

```
cwnd <= max_ssthresh
```

cwnd is increased by one MSS (MAXIMUM SEGMENT SIZE) for every arriving ACK (acknowledgement) during slow-start, as is always the case. For:

```
cwnd > max_ssthresh
```

the invariant is maintained that the congestion window is increased during slow-start by at most max_ssthresh/2 MSS per round-trip time.

This is done as follows:

```
For each arriving ACK in slow-start:
  If (cwnd <= max_ssthresh)
     cwnd += MSS;
  else
     K = int(cwnd/(0.5 max_ssthresh));
     cwnd += int(MSS/K);
```

Thus during Limited Slow-Start the window is increased by 1/K MSS for each arriving ACK, for K = int(cwnd/(0.5 max_ssthresh)), instead of by 1 MSS as in the standard slow-start [RFC2581].

Our current recommendation would be for max_ssthresh to be set to 100 MSS. (This is illustrated in the NS simulator, for snapshots after May 1, 2002, in the tests "./test-all-tcpHighspeed tcp1A" and "./test-all-tcpHighspeed tcpHighspeed1" in the subdirectory "tcl/lib".)

With Limited Slow-Start, when the congestion window is greater than max_ssthresh the window is increased by at most 1/2 MSS for each arriving ACK, when the congestion window is greater than 1.5 max_ssthresh the window is increased by at most 1/3 MSS for each arriving ACK, and so on.

With Limited Slow-Start it takes:

```
log(max_ssthresh)
```

round-trip times to reach a congestion window of max_ssthresh, and it takes:

```
log(max_ssthresh) + (cwnd - max_ssthresh)/(max_ssthresh/2)
```

round-trip times to reach a congestion window of cwnd, for a congestion window greater than max_ssthresh.

By limiting the maximum increase in the congestion window in a round-trip time, Limited Slow-Start can reduce the number of drops during slow-start, and improve the performance of TCP connections with large congestion windows.

3. Related Proposals

There has been considerable research on mechanisms for the TCP sender to learn about the limitations of the available bandwidth, and to exit slow-start before receiving a congestion indication from the network [VEGAS,H96]. This draft proposes a simple limitation on slow-start that can be effective in some cases even in the absence of such mechanisms.

Rate-based pacing has also been proposed to improve the performance of TCP during slow-start [VH97,AD98,KCRP99,ASA00]. We believe that rate-based pacing could be of significant benefit, and could be used in addition to the Limited Slow-Start in this proposal.

Appropriate Byte Counting [A01] proposes that TCP increase its congestion window as a function of the number of bytes acknowledged, rather than as a function of the number of ACKs received. Appropriate Byte Counting is largely orthogonal to this proposal for Limited Slow-Start.

(If I have missed acknowledgements to other relevent proposals, including to other proposals to set the initial value of ssthresh, please let me know.)

4. Acknowledgements

This proposal is part of a larger proposal for HighSpeed TCP for TCP connections with large congestion windows, and resulted from simulations done by Evandro de Souza, in joint work with Deb Agarwal. This proposal for Limited Slow-Start drew in part from discussions with Tom Kelly, who has used a similar modified slow-start in his own research with congestion control for high-bandwidth connections.

5. Normative References

[A01] Mark Allman, "TCP Congestion Control with Appropriate Byte Counting", draft-allman-tcp-abc-02.txt, internet-draft, work-in-progress, November 2001.

[RFC2581] M. Allman and V. Paxson, "TCP Congestion Control", RFC 2581, April 1999.

6. Informative References

[AD98] Mohit Aron and Peter Druschel, "TCP: Improving Start-up Dynamics by Adaptive Timers and Congestion Control"", TR98-318, Rice University, 1998. URL "http://cs-tr.cs.rice.edu/Dienst/UI/2.0/Describe/ncstrl.rice_cs/TR98-318/".

[ASA00] A. Aggarwal, S. Savage, and T. Anderson, "Understanding the Performance of TCP Pacing", Proceedings of the 2000 IEEE Infocom Conference, Tel-Aviv, Israel, March, 2000. URL "http://www.cs.ucsd.edu/˜savage/".

[F02] S. Floyd, "Performance Problems with TCP's Slow-Start", 2002. URL "http://www.icir.org/floyd/hstcp/slowstart/".

[H96] J. C. Hoe, "Improving the Start-up Behavior of a Congestion Control Scheme for TCP", SIGCOMM 96, 1996. URL "http://www.acm.org/sigcomm/sigcomm96/program.html".

[KCRP99] J. Kulik, R. Coulter, D. Rockwell, and C. Partridge, "A Simulation Study of Paced TCP", BBN Technical Memorandum No. 1218, 1999. URL "http://mimas.lcs.mit.edu/~jokulik/tcppacing.html".

[VEGAS] Vegas Web Page, University of Arizona. URL "http://www.cs.arizona.edu/protocols/".

[VH97] Vikram Visweswaraiah and John Heidemann, "Rate Based Pacing for TCP", 1997. URL "http://www.isi.edu/lsam/publications/rate_based_pacing/".

7. Security Considerations

This proposal makes no changes to the underlying security of TCP.

8. IANA Considerations

There are no IANA considerations regarding this document.

AUTHORS' ADDRESSES

Sally Floyd
Phone: +1 (510) 666-2989
ICIR (ICSI Center for Internet Research)
Email: floyd@icir.org
URL: http://www.icir.org/floyd/

This draft was created in June 2002.